Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
C. R. Acad. Sci. Paris, Ser. I 340 (2005) 525–528
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistique/Probabilités

Vitesses de convergence uniforme presque sûre d'estimateurs non-paramétriques de la régression

David Blondin, Anne Massiani, Pierre Ribereau

L.S.T.A., boîte 158, 175, rue du Chevaleret, 75013 Paris, France

Reçu le 4 septembre 2004 ; accepté après révision le 1er février 2005

Disponible sur Internet le 5 mars 2005

Présenté par Paul Deheuvels

Résumé

Nous établissons des vitesses de convergence uniforme presque sûre d'estimateurs non-paramétriques de la fonction de régression, tels que l'estimateur localement linéaire et certains estimateurs par la méthode des ondelettes. Notre technique de démonstration s'appuie sur une méthodologie récente développée par Deheuvels et Mason (Statist. Inference Stoch. Process. 7 (3) (2004) 225–277), fondée sur la théorie des processus empiriques indexés par des classes de fonctions. ***Pour citer cet article : D. Blondin et al., C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

Abstract

Rates of strong consistency for nonparametric regression estimators. We obtain rates of strong uniform consistency for some nonparametric regression estimators, including the local linear regression and some wavelet estimators. Our method of proof relies on recent empirical process theory developed by Deheuvels and Mason (Statist. Inference Stoch. Process. 7 (3) (2004) 225–277). ***To cite this article: D. Blondin et al., C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

1. Introduction

Soit (X, Y) un couple de variables aléatoires réelles, de densité f_{XY} sur $\mathbb{R}^2$ et de lois marginales de densités f_X et f_Y. Nous nous intéressons ici à l'estimation de la fonction de régression $m(\cdot) = \mathbb{E}[Y|X = \cdot]$ à partir d'une suite d'observations $(X_1, Y_1), (X_2, Y_2), \ldots,$ indépendantes et de même loi que (X, Y). En s'appuyant sur l'étude du processus empirique local indexé par certaines classes de fonctions, Deheuvels et Mason [2] ainsi que Einmahl et Mason [4] ont développé récemment une méthodologie qui permet d'établir des vitesses de convergence uniforme

Adresses e-mail : blondin@ccr.jussieu.fr (D. Blondin), amassia@ccr.jussieu.fr (A. Massiani), ribereau@ccr.jussieu.fr (P. Ribereau).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2005.02.008

presque sûre pour la déviation stochastique de l'estimateur non-paramétrique de la régression de Nadaraya–Watson (cf. [7,8]). Le but de cette Note est d'étendre ces résultats à des estimateurs non-paramétriques de la régression plus performants. Nous considérons dans un premier temps l'estimateur localement linéaire (cf. (1), ci-dessous), étudié en détails par Fan [5,6], qui s'adapte à tous types de dispositifs expérimentaux (fixes ou aléatoires) ainsi qu'à l'estimation aux bords. Cet estimateur a de plus un meilleur biais que l'estimateur de Nadaraya–Watson et est minimax linéaire sur la classe des fonctions de régression à dérivée seconde bornée. Nous nous intéressons ensuite à l'estimateur linéaire par la méthode des ondelettes considéré par Antoniadis, Grégoire et McKeague [1], puis à l'estimateur à seuil étudié par Delyon et Juditsky [3]. Nous obtenons de nouvelles lois limites pour la convergence uniforme presque sûre de ces estimateurs. Les résultats sont présentés dans les Sections 2 et 3. Quelques éléments de preuves sont donnés dans la Section 4.

2. Estimateur localement linéaire

Soit K un noyau à support compact inclus dans un intervalle borné I, continu sur I, à variation bornée et tel que $\int_{-\infty}^{\infty} K(s)\,\mathrm{d}s = 1$. Soit $\{h_n,\ n \geqslant 1\}$ une suite de réels strictement positifs, telle que $h_n \to 0$ lorsque $n \to \infty$. L'estimateur localement linéaire est alors défini, pour tout $x \in \mathbb{R}$ et tout $n \geqslant 1$, par :

$$\widehat{m}_n^{LL}(x) = \frac{\hat{r}_{n,0}(x)\hat{f}_{n,2}(x) - \hat{r}_{n,1}(x)\hat{f}_{n,1}(x)}{\hat{f}_{n,2}(x)\hat{f}_{n,0}(x) - \{\hat{f}_{n,1}(x)\}^2}, \tag{1}$$

où on a posé, pour $j \in \{0, 1, 2\}$,

$$\hat{f}_{n,j}(x) = \frac{1}{nh_n}\sum_{i=1}^{n}\left(\frac{x - X_i}{h_n}\right)^j K\left(\frac{x - X_i}{h_n}\right) \quad \text{et} \quad \hat{r}_{n,j}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} Y_i\left(\frac{x - X_i}{h_n}\right)^j K\left(\frac{x - X_i}{h_n}\right).$$

Par convenance, on note, pour $j \in \{0, 1, 2\}$,

$$f_{n,j}(x) = \mathbb{E}\big[\hat{f}_{n,j}(x)\big] \quad \text{et} \quad r_{n,j}(x) = \big[\mathbb{E}\hat{r}_{n,j}(x)\big].$$

Nous obtenons ici un contrôle de la déviation maximale de l'estimateur $\widehat{m}_n^{LL}$ par rapport au terme de centrage suivant :

$$m_n^{LL}(x) = \frac{r_{n,0}(x)f_{n,2}(x) - r_{n,1}(x)f_{n,1}(x)}{f_{n,0}(x)f_{n,2}(x) - \{f_{n,1}(x)\}^2}, \quad \text{pour } x \in \mathbb{R},\ n \geqslant 1.$$

Théorème 2.1. *Soit $[A, B]$ un intervalle de $\mathbb{R}$. Supposons que les conditions suivantes sur la distribution du couple (X, Y) soient vérifiées* :

– f_{XY} *est continue sur* $[A, B] \times \mathbb{R}$, (2)

– f_X *est continue et strictement positive sur* $[A, B]$, (3)

– Y *est bornée*. (4)

Si $\{h_n, n \geqslant 1\}$ *est une suite qui vérifie les conditions* :

$$h_n \downarrow 0, \quad nh_n \uparrow \infty, \quad \frac{\log(h_n^{-1})}{\log\log n} \to \infty, \quad \frac{nh_n}{\log n} \to \infty, \quad \textit{lorsque } n \to \infty, \tag{5}$$

alors, pour tout $A < C < D < B$, on a, presque sûrement :

$$\lim_{n\to\infty}\left\{\frac{nh_n}{2\log(1/h_n)}\right\}^{1/2} \sup_{C\leqslant x\leqslant D} \pm\big\{\widehat{m}_n^{LL}(x) - m_n^{LL}(x)\big\} = \sup_{C\leqslant x\leqslant D}\left\{\frac{\mathrm{Var}[Y|X=x]}{f_X(x)}\int_{\mathbb{R}} K^2(u)\,\mathrm{d}u\right\}^{1/2}. \tag{6}$$

Remarque 1. Il est possible d'obtenir des lois limites similaires à (6) dans le cadre plus général de l'estimation localement polynomiale d'ordre p de la dérivée k-ième de la régression, avec $(p,k)\in\mathbb{N}^2$ et $k<p$. Ces résultats ne seront pas exposés dans cette Note, par souci de concision.

3. Estimateurs par la méthode des ondelettes

Soient φ et Ψ une fonction d'échelle et une ondelette mère supposées à support compact. Notons pour tout $j,k\in\mathbb{Z}$, $\varphi_{jk}(\cdot)=2^{j/2}\varphi(2^j\cdot-k)$, $\Psi_{jk}(\cdot)=2^{j/2}\Psi(2^j\cdot-k)$, et désignons par K le noyau généralisé associé à φ défini, pour tout $x,y\in\mathbb{R}$, par $K(x,y)=\sum_{k\in\mathbb{Z}}\varphi(x-k)\varphi(y-k)$. Soient $j_0\in\mathbb{Z}$ et j_n un niveau d'analyse multi-résolution vérifiant $j_n\to\infty$ quand $n\to\infty$. Nous nous intéressons tout d'abord à l'estimateur linéaire considéré par Antoniadis, Gregoire et McKeague [1], défini par :

$$\widehat{m}_n=\left\{\sum_{k\in\mathbb{Z}}\hat{\alpha}_{j_0k}\varphi_{j_0k}+\sum_{j=j_0}^{j_n}\hat{\beta}_{jk}\Psi_{jk}\right\}\Big/\hat{f}_n,$$

où

$$\hat{\alpha}_{j_0k}=\frac{1}{n}\sum_{i=1}^{n}\varphi_{j_0k}(X_i),\qquad \hat{\beta}_{jk}=\frac{1}{n}\sum_{i=1}^{n}Y_i\Psi_{jk}(X_i),\qquad \hat{f}_n=\frac{2^{j_n}}{n}\sum_{i=1}^{n}K(2^{j_n}\cdot,2^{j_n}X_i).$$

Nous étudions dans le théorème ci-dessous la vitesse de convergence uniforme presque sûre de l'erreur stochastique $\hat{m}_n(x)-r_n(x)/f_n(x)$, où on a noté,

$$r_n=\mathbb{E}\left[\sum_{k\in\mathbb{Z}}\hat{\alpha}_{j_0k}\varphi_{j_0k}+\sum_{j=j_0}^{j_n}\hat{\beta}_{jk}\Psi_{jk}\right],\quad \text{et}\quad f_n=\mathbb{E}[\hat{f}_n],$$

la différence $r_n(x)/f_n(x)-m(x)$ pouvant être contrôlée par des méthodes analytiques.

Théorème 3.1. *Supposons que la fonction φ soit continue à gauche, à variation bornée, et à support compact, et que les conditions* (2), (3) *et* (4) *soient vérifiées. Si de plus $j_n=[l_n]$, où $\{h_n=2^{-l_n},\ n\geqslant1\}$ est une suite satisfaisant* (5), *alors, pour tout $A<C<D<B$, on a, presque sûrement* :

$$\lim_{n\to\infty}\left\{\frac{n}{2^{(j_n+1)}\log(2^{j_n})}\right\}^{1/2}\sup_{C\leqslant x\leqslant D}\left\{\frac{\pm\{\widehat{m}_n(x)-r_n(x)/f_n(x)\}}{\{\int_{-\infty}^{\infty}K^2(2^{j_n}x,u)\,\mathrm{d}u\}^{1/2}}\right\}=\sup_{C\leqslant x\leqslant D}\left\{\frac{\operatorname{Var}[Y|X=x]}{f_X(x)}\right\}^{1/2}.$$

Ce premier résultat ouvre la voie à l'étude d'autres estimateurs plus sophistiqués tels que les estimateurs à seuil. Dans le cas où la variable X suit une loi uniforme sur $[0,1]$, la densité f_X n'ayant plus à être estimée, Delyon et Juditsky [3] ont considéré l'estimateur $\widehat{m}_n^*$, défini par,

$$\widehat{m}_n^*=\sum_{k\in\mathbb{Z}}\hat{\alpha}_{j_0k}\varphi_{j_0k}+\sum_{j=j_0}^{j_n}\hat{\beta}_{jk}\mathbb{1}_{\{|\hat{\beta}_{jk}|>c\sqrt{j/n}\}}\Psi_{jk},\tag{7}$$

où c est une constante strictement positive et $\{j_n,n\geqslant1\}$ désigne une suite telle que $2^{j_n}\simeq n/\log n$. Ils ont montré que cet estimateur est adaptatif pour l'erreur L_p sur un large ensemble de classes de fonctions. Nous étudions dans le corollaire suivant sa vitesse de convergence uniforme presque sûre.

Corollaire 3.2. *Supposons que φ vérifie les hypothèses du Théorème* 3.1, *et qu'il existe un entier $N\geqslant0$ tel que* :

$$\int_{-\infty}^{\infty}(y-x)^kK(x,y)\,\mathrm{d}y=\mathbb{1}_{\{k=0\}},\quad \textit{pour tout } k=0,\dots,N,\ x\in\mathbb{R}.$$

Supposons également que $m(\cdot)$ appartienne à l'espace de Besov B_p^{sq}, où $1 \leqslant p, q \leqslant \infty$ et $1/p < s < N+1$, et que les conditions (2) *et* (4) *soient satisfaites. Si la constante c intervenant dans* (7) *est choisie telle que $c^2 > 4[\|m\|_\infty + c/3\|\Psi\|_\infty]$, alors, pour tout $A < C < D < B$, on a, presque sûrement* :

$$\sup_{C \leqslant x \leqslant D} \left|\widehat{m}_n^*(x) - m(x)\right| = \mathrm{O}\left(\frac{\log n}{n}\right)^\beta, \quad \text{où } \beta = \frac{s - 1/p}{2s + 1 - 2/p}. \tag{8}$$

Remarque 2. Notons que la vitesse de convergence uniforme presque sûre obtenue dans (8) est la même que celle du risque uniforme $\mathbb{E}[\|\widehat{m}_n^*(x) - m(x)\|_\infty]$ établie par Delyon et Juditsky [3].

4. Eléments de preuve

Par souci de concision, nous ne traitons ici que le cas de l'estimateur localement linéaire. L'idée principale de la preuve du Théorème 2.1 consiste à ramener notre problème à l'étude d'une version linéarisée de la déviation $\hat{m}_n^{LL} - m_n^{LL}$, qui s'exprime en fonction des quantités $\hat{f}_{n,j} - f_{n,j}$, et $\hat{r}_{n,j} - r_{n,j}$, pour $j \in \{0, 1, 2\}$. Il suffit alors pour terminer la démonstration d'utiliser la méthodologie de Deheuvels et Mason [2], qui permet d'étudier ces déviations. Leur comportement asymptotique presque sûr peut être analysé via le processus ci-dessous, défini, pour tout $x \in \mathbb{R}$ et $j \in \{0, 1, 2\}$, par

$$W_{n,j}(x) := \sum_{i=1}^n \bigl(c(x)Y_i + d(x)\bigr)\left(\frac{x - X_i}{h_n}\right)^j K\left(\frac{x - X_i}{h_n}\right) - n\mathbb{E}\left\{\bigl(c(x)Y + d(x)\bigr)\left(\frac{x - X}{h_n}\right)^j K\left(\frac{x - X}{h_n}\right)\right\},$$

où $c(x)$ et $d(x)$ désignent deux fonctions auxiliaires supposées continues sur l'intervalle $[A, B]$.

Références

[1] A. Antoniadis, G. Gregoire, I.W. McKeague, Wavelet methods for curve estimation, J. Amer. Statist. Assoc. 89 (1994) 1340–1353.
[2] P. Deheuvels, D.M. Mason, General asymptotic confidence bands based on kernel-type function estimators, Statist. Inference Stoch. Process. 7 (3) (2004) 225–277.
[3] B. Delyon, A. Juditsky, On minimax wavelet estimators, Appl. Comput. Harmon. Anal. 3 (1996) 215–228.
[4] U. Einmahl, D. Mason, An empirical process approach to the uniform consistency of kernel-type function estimators, J. Theoret. Probab. 13 (2000) 1–37.
[5] J. Fan, Design-adaptative nonparametric regression, J. Amer. Statist. Assoc. 87 (1992) 998–1004.
[6] J. Fan, Local linear regression smoothers and their minimax efficiencies, Ann. Statist. 21 (1992) 196–216.
[7] E.A. Nadaraya, On estimating regression, Theoret. Probab. Appl. 9 (1964) 141–142.
[8] G.S. Watson, Smooth regression analysis, Sankhyà Ser. A 26 (1964) 359–372.